ITERATIVELY REWEIGHTED FGMRES AND FLSQR FOR SPARSE RECONSTRUCTION*

SILVIA GAZZOLA †, JAMES G. NAGY ‡, AND MALENA SABATÉ LANDMAN§

Abstract. This paper presents two new algorithms to compute sparse solutions of large-scale linear discrete ill-posed problems. The proposed approach consists in constructing a sequence of quadratic problems approximating an ℓ_2-ℓ_1 regularization scheme (with additional smoothing to ensure differentiability at the origin) and partially solving each problem in the sequence using flexible Krylov-Tikhonov methods. These algorithms are built upon a new solid theoretical justification that guarantees that the sequence of approximate solutions to each problem in the sequence converges to the solution of the considered modified version of the ℓ_2-ℓ_1 problem. Compared to other traditional methods, the new algorithms have the advantage of building a single (flexible) approximation (Krylov) subspace that encodes regularization through variable "preconditioning" and that is expanded as soon as a new problem in the sequence is defined. Links between the new solvers and other well-established solvers based on augmenting Krylov subspaces are also established. The performance of these algorithms is shown through a variety of numerical examples modeling image deblurring and computed tomography.

Key words. Krylov Methods, Inverse Problems, Sparse reconstruction, Flexible GMRES, Flexible LSQR, augmented Krylov methods, Image Deblurring, Computed Tomography

AMS subject classifications. 65F20, 65F22, 65F30

1. Introduction.

Large-scale linear ill-posed inverse problems of the form

$$A x_{true} = b_{true} + e = b, \quad A \in \mathbb{R}^{m\times n}, \tag{1.1}$$

where x_{true} is the desired unknown solution and e is some unknown Gaussian white noise that affects the data b, arise in the discretization of problems stemming from various scientific and engineering applications, such as astronomical and biomedical imaging, or computed tomography in medicine and industry. In particular, we are interested in the case where A is ill-conditioned with ill-determined rank, i.e., the singular values of A decay and cluster at zero without an evident gap between two consecutive ones to indicate numerical rank. In this case, due to the presence of noise in the measured data, the naive solution $A^\dagger b$ of (1.1) (where $A^\dagger$ is the Moore-Penrose pseudoinverse of A) can be very different from the desired solution, $A^\dagger b_{true}$, due to noise amplification; see, e.g., [23]. Therefore, to obtain a meaningful approximation of x_{true}, problem (1.1) should be regularized, i.e., replaced by a closely related problem whose solution is less sensitive to perturbations in the data b (for a more detailed discussion on ill-posed and discrete ill-posed problems and regularization see, e.g., [25]).

One of the most well-known approaches for regularizing linear ill-posed problems is Tikhonov regularization, which, in its general formulation, computes a regularized

*Submitted to the editors DATE.

Funding: S. Gazzola is partially funded by EPSRC, under grant EP/T001593/1. J. Nagy is partially supported by U.S. National Science Foundation under Grant DMS-1819042. M. Sabaté Landman is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under project EP/L015684/1.

†Department of Mathematical Sciences, University of Bath, United Kingdom (S.Gazzola@bath.ac.uk, https://people.bath.ac.uk/sg968/).

‡Department of Mathematics, Emory University, Atlanta (jnagy@emory.edu).

§Department of Mathematical Sciences, University of Bath, United Kingdom (M.Sabate.Landman@bath.ac.uk, https://people.bath.ac.uk/msl39/)

approximation to the solution of (1.1) by solving the following minimization problem

$$x_{\lambda,L} = \min_x \|Ax - b\|_2^2 + \lambda \|Lx\|_2^2 \,. \tag{1.2}$$

Here, the regularization parameter $\lambda > 0$ balances the effect of the fit-to-data term $\|Ax - b\|_2^2$ and the regularization term $\|Lx\|_2^2$. The regularization matrix $L \in \mathbb{R}^{q\times n}$ has the effect of enhancing certain properties on the solution and it is usually chosen to be the identity (in this case, problem (1.2) is said to be in standard form) or a rescaled finite differences approximation of a derivative operator (to enforce smoother solutions); if the null space of A and the null space of L intersect trivially, the general-form Tikhonov solution $x_{\lambda,L}$ is unique.

For large-scale problems, where A does not have an exploitable structure nor is even explicitly stored (i.e., may be defined as a function that efficiently computes the actions of A and, possibly, A^T, on vectors), the only way to solve problem (1.1) is to apply an iterative method to obtain a sequence of approximated solutions $\{x_k\}_{k\geq 1}$. In fact, many well-known general iterative solvers, e.g., Landweber and Kaczmarz methods, and many Krylov subspace methods, leverage the so-called "semiconvergence" phenomenon and lead to a regularized solution if the iterations are stopped sufficiently early, with the number of iterations playing the role of a discrete regularization parameter (see [25, Chapter 6] for a more accurate description). This paper will only consider the GMRES and LSQR iterative methods, and variations thereof: these are Krylov methods that compute a regularized solution by expanding an approximation subspace for the solution and solving a projected least squares problem at each iteration. Note that LSQR is mathematically equivalent to CGLS.

When regularization relies on semiconvergence only, a bad stopping criterion can lead to a big error in the approximated solution. Moreover, semiconvergence may happen before the relevant basis vectors for the solution are incorporated in the Krylov approximation subspace for the solution; see [25, Chapter 6] and [28] for more details. These issues can be mitigated by applying further regularization within the iterations, e.g., by using schemes that combine an iterative Krylov solver and Tikhonov regularization, as detailed below. Consider, for simplicity, $L = I$ in (1.2), i.e., Tikhonov regularization in standard form. Projecting (1.2) into a kth dimensional Krylov subspace spanned by the columns of the matrix V_k leads to

$$x_k = V_k y_k, \quad y_k = \arg\min_y \|AV_k y - b\|_2^2 + \lambda \|V_k y\|_2^2, \tag{1.3}$$

which is sometimes referred to as "first-regularize-then-project" approach [25, Chapter 6]. Alternatively, a "first-project-then-regularize" approach can also be used, which involves projecting the original linear system (1.1) and then applying standard Tikhonov regularization, leading to

$$x_k = V_k y_k, \quad y_k = \arg\min_y \|AV_k y - b\|_2^2 + \lambda \|y\|_2^2. \tag{1.4}$$

For fixed λ, and assuming the columns of V_k to be orthonormal, expressions (1.3) and (1.4) are equivalent and both schemes are interchangeable. Methods employing the latter approach are also known as hybrid methods [11, 37] and they have recently attracted a lot of attention in the case of large-scale problems where the regularization parameter λ is not known a priori; see [10, 19, 21, 30]. Indeed, hybrid methods allow for a very efficient (local) choice of the parameter $\lambda = \lambda_k$ at each iteration $k \ll \min\{m, n\}$; moreover, when k increases, λ_k seems to stabilize around a value that is suitable for the full-dimensional problem (1.2).

Tikhonov regularization as defined in (1.2) is rather restrictive, and more general regularization strategies can yield to better approximations of the solution of (1.1). In particular, this paper focuses on regularized problems of the form

$$\min_x \|Ax - b\|_2^2 + \lambda \|x\|_p^p\,, \tag{1.5}$$

where, for $0 < p \leq 1$, the ℓ_p-norm regularization term enforces sparsity in the solution. Although sparse vectors have a small ℓ_0 "norm", considering an ℓ_0 regularization term yields to an NP hard optimization problem (1.5); see [16]. Therefore, it is common to approximate the ℓ_0 regularization term by an ℓ_p term with $0 < p \leq 1$, noting that for $0 < p < 1$ problem (1.5) is nonconvex, and for $p = 1$ problem (1.5) approximates the desired ℓ_0-norm via convex relaxation but is non-differentiable at the origin; see, e.g., [27, 31, 32]. Note that, if sparsity of the solution is assumed in a different domain (e.g., wavelets or discrete cosine transform) a sparsity transform can be incorporated in the regularization term. The values $0 < p \leq 2$ will be considered in this paper; when $p = 2$, problem (1.5) reduces to Tikhonov regularization in standard form. The ℓ_2-ℓ_p regularization problem (1.5) can be solved by a variety of optimization methods [4, 22, 33, 46], or by employing iterative schemes that approximate the regularization term in (1.5) by a sequence of weighted ℓ_2 terms [39]. Methods of the second kind come equipped with (local) convergence proofs for most values of $p > 0$, but usually rely on inner-outer schemes so they can become very expensive computationally; see, e.g., [5, Chapter 4].

More recently, solvers for the ℓ_2-ℓ_p regularization problem that avoid nested loops of iterations by combining reweighting techniques and modified Krylov methods have gained popularity. Namely, generalized Krylov subspaces are considered in [31, 27, 6], and hybrid solvers based on the flexible Arnoldi and the flexible Golub-Kahan decompositions are considered in [9, 18, 20].

In this paper, we propose two new iterative Krylov-Tikhonov methods that use the flexible Arnoldi and the flexible Golub-Kahan decomposition, respectively, to solve the ℓ_2-ℓ_p regularization problem (1.5) by building a single approximation subspace through the iterations. Both algorithms are essentially different from the strategies already available in the literature. On the one hand, differently from [31, 27, 6], the approach proposed in this paper is based on flexible Krylov subspaces. On the other hand, differently from the "first-project-then-regularize" scheme corresponding to hybrid methods implicitly adopted in [9, 18], the approach proposed in this paper exploits a "first-regularize-then-project" scheme. In fact, another contribution of this paper is to show that regularizing and projecting are not interchangeable anymore in the flexible Krylov subspace setting, and properties derived from using the "first-regularize-then-project" approach are used to provide theoretical justification of convergence for the newly proposed algorithms. An original interpretation of the new algorithms in the general framework of augmented and recycled Krylov subspaces is also given. It should be stressed that both the new algorithms are inherently "matrix-free" (i.e., they only require the action of A on vectors, and additionally the action of A^T if the flexible Golub-Kahan decomposition is considered), and allow for an iteration dependent choice of the regularization parameter.

The paper is organized as follows. In Section 2 background material on ℓ_2-ℓ_p regularization is reviewed. In particular, Section 2 explains how to approximate the ℓ_p regularization term in (1.5) using an iteratively reweighted scheme, and how the transformation of the resulting problem into standard form leads to iteration-dependent right preconditioning for a Tikhonov problem of the form (1.2). In Section 3 two new algorithms for sparse reconstruction (called IRW-FGMRES and IRW-FLSQR) are introduced, along with a solid theoretical proof of convergence and links with augmented Krylov subspace methods. Finally, numerical results are presented in Section 4, and general conclusions are given in Section 5.

2. Background on ℓ_2-ℓ_p regularization.

Iteratively reweighted schemes for the ℓ_2-ℓ_p regularization problem intrinsically rely on the interpretation of problem (1.5) as a non-linear weighted least squares problem of the form

$$\min_x \|Ax-b\|_2^2 + \lambda\|x\|_p^p = \min_x \|Ax-b\|_2^2 + \lambda\|W^{(p)}(x)x\|_2^2, \tag{2.1}$$

where the diagonal weighting $W^{(p)}(x)$ is defined as

$$W^{(p)}(x) = \operatorname{diag}\left((|[x]_i|^{\frac{p-2}{2}})_{i=1,\dots,n}\right), \tag{2.2}$$

and $[x]_i$ denotes the ith component of the vector x. Note that, when $0 < p < 2$, division by zero might occur if $[x]_i = 0$ for any $i \in \{1, ..., n\}$ and, in fact, this is a far from unlikely situation in the case of sparse solutions. For this reason, in this paper, instead of (2.2), the following closely related weights are considered

$$\widetilde{W}^{(p,\tau)}(x) = \operatorname{diag}\left((([x]_i^2+\tau^2)^{\frac{p-2}{4}})_{i=1,\dots,n}\right), \tag{2.3}$$

where τ is a fixed parameter chosen ahead of the iterations, and problem (2.1) is replaced by

$$\min_x \underbrace{\|Ax-b\|_2^2 + \lambda\|\widetilde{W}^{(p,\tau)}(x)x\|_2^2}_{T^{(p,\tau)}(x)}, \tag{2.4}$$

where $\tau \neq 0$ also ensures that $T^{(p,\tau)}(x)$ is differentiable at the origin for $p > 0$. Note that (2.4) should be considered a smooth version of problem (2.1) and, formally, problem (2.1) can be recovered from problem (2.4) setting $\tau = 0$.

A well established framework to solve problem (2.4) is the local approximation of $T^{(p,\tau)}$ by a sequence of quadratic functionals $T_k(x)$ that give rise to a sequence of quadratic problems of the form

$$x_{k,\star} = \arg\min_x \underbrace{\|Ax-b\|_2^2 + \lambda\|W_k x\|_2^2 + c_k}_{T_k(x)}, \tag{2.5}$$

where $W_k = \widetilde{W}^{(p,\tau)}(x_{k-1,\star})$. Here, c_k (a constant term for the kth problem in the sequence with respect to x), and λ (which has absorbed other possible multiplicative constants with respect to (2.4)) are chosen so that $T_k(x)$ in (2.5) corresponds to a quadratic tangent majorant of $T^{(p,\tau)}(x)$ in (2.4) at $x = x_{k-1,\star}$. By definition, this implies that $T_k(x) \geq T^{(p,\tau)}(x)$ for all $x \in \mathbb{R}^n$, $T_k(x_{k-1,\star}) = T^{(p,\tau)}(x_{k-1,\star})$, and $\nabla T_k(x_{k-1,\star}) = \nabla T^{(p,\tau)}(x_{k-1,\star})$; see also [27, 39]. Since p and τ are chosen ahead of the iterations, they are omitted from the notations for the weighting matrix W_k.

The vector $x_{k,\star}$ formally denotes the solution of (2.5). For moderate-scale problems, or for large-scale problems where A has some exploitable structure, $x_{k,\star}$ may be obtained by applying a direct solver to (2.5). For large-scale unstructured problems, only iterative solvers can be used in different fashions to approximate the solution of (2.5), naturally leading to an inner-outer iteration scheme for the sequence of problems

(2.4). This is the case considered in the present paper, so that $x_{k,\star}$ corresponds to the approximate solution $x_{k,l}$ of the kth problem of the form (2.5) (or 'at the kth outer iteration') at the lth iteration of the inner cycle of iterations. Iteratively Reweighted Least Squares (IRLS) or Iteratively Reweighted Norm (IRN) methods based on an inner-outer iteration scheme are very popular [12, 39] and have been used in combination with different inner solvers, such as steepest descent and CGLS. Typically $x_{k,\star} = x_{k,l}$ is obtained when a stopping criterion is satisfied for problem (2.5) to indicate convergence of the approximate solution; alternatively, problem (2.5) can be partially solved and $x_{k,\star} = x_{k,l}$ denotes the latest available approximation of x. In any case, $T_k(x)$ in (2.5) is a quadratic tangent majorant of $T^{(p,\tau)}(x)$ in (2.4) at $x = x_{k-1,\star}$, and IRLS or IRN approaches are particular instances of majorization-minimization (MM) schemes: for fixed λ, it is known that solving a sequence of problems of the form (2.5) produces a sequence of approximate solutions that converge to the minimizer of problem (2.4); see, e.g., [12]. Fully solving each problem (2.5) can result in a computationally demanding scheme.

For W_k square and invertible (note that this can be assumed for any fixed $p > 0$ when the weights are defined as in (2.3) with $\tau > 0$), problem (2.5) can be easily and conveniently transformed into standard form as follows

$$\bar{x}_{k,\star} = \arg\min_{\bar{x}} \|AW_k^{-1}\bar{x} - b\|_2^2 + \lambda\|\bar{x}\|_2^2\,, \quad \text{so that} \quad x_{k,\star} = W_k^{-1}\bar{x}_{k,\star}. \tag{2.6}$$

The interpretation of the matrix W_k^{-1} as a right preconditioner for problem (2.5) can be exploited under the framework of prior-conditioning [7]. The simplest way to use formulation (2.6) in combination with Krylov methods is to rely on an inner-outer scheme (e.g., with an inner loop of (hybrid) GMRES or LSQR iterations [9, 18]) so that, at each outer iteration, a new Krylov subspaces is built. Let $V_{k,l} \in \mathbb{R}^{n\times l}$ be the matrix whose columns, at the lth inner iteration of the kth outer cycle, span a Krylov subspace $\mathcal{K}_{k,l}$ of dimension l. Then, problem (2.6) can be projected and solved in $\mathcal{K}_{k,l}$ by computing

$$\bar{y}_{k,l} = \arg\min_{\bar{y}} \|A\overbrace{W_k^{-1}\underbrace{V_{k,l}\bar{y}}_{\bar{x}}}^{x} - b\|_2^2 + \lambda\|\underbrace{V_{k,l}\bar{y}}_{\bar{x}}\|_2^2\,, \tag{2.7}$$

so that $\bar{x}_{k,l} = V_{k,l}\,\bar{y}_{k,l}$ and $x_{k,l} = W_k^{-1}\,\bar{x}_{k,l} = W_k^{-1}\,V_{k,l}\,\bar{y}_{k,l}$. Note that, since $V_{k,l}$ has orthonormal columns, solving equation (2.7) is equivalent to solving

$$\bar{y}_{k,l} = \arg\min_{\bar{y}} \|A\overbrace{\underbrace{W_k^{-1}V_{k,l}}_{Z_{k,l}}\,\bar{y}}^{x} - b\|_2^2 + \lambda\|\bar{y}\|_2^2, \tag{2.8}$$

which is consistent with the idea of "first-regularize-then-project" being equivalent to "first-project-then-regularize" for hybrid solvers (cf. [25, Chapter 6]). An alternative interpretation of this scheme is that, at the lth inner iteration of the kth outer cycle, an approximate solution to the original problem is sought in the preconditioned space $\mathcal{R}(Z_{k,l}) = \mathcal{R}(W_k^{-1}V_{k,l})$, where $\mathcal{R}(\cdot)$ denotes the range of a matrix. Note that, when applying preconditioned GMRES,

$$\begin{aligned}\mathcal{R}(Z_{k,l}) &= W_k^{-1}\mathcal{K}_l(AW_k^{-1}, b)\\ &= \operatorname{span}\{W_k^{-1}b, W_k^{-1}(AW_k^{-1})b, ..., W_k^{-1}(AW_k^{-1})^{l-1}b\}\,,\end{aligned} \tag{2.9}$$

while, when applying preconditioned LSQR,

$$\begin{aligned}(2.10) \qquad \mathcal{R}(Z_{k,l}) &= W_k^{-1}\mathcal{K}_l(W_k^{-1}A^TAW_k^{-1}, W_k^{-1}A^Tb) \\ &= \text{span}\{(W_k^{-1})^2A^Tb, , ..., ((W_k^{-1})^2A^TA)^{l-1}(W_k^{-1})^2A^Tb\}\,.\end{aligned}$$

With respect to preconditioned GMRES, preconditioned LSQR naturally applies the inverse of the weight matrix W_k twice for every new direction included in the search space, and hence, twice at each iteration.

It should be stressed that, for both (2.7) and (2.8) to be equivalent to (2.6), the regularization term in (2.7) has to be $\|V_{k,l}\bar{y}\|_2^2$, where $V_{k,l}\,\bar{y} = \bar{x}$ in (2.6), and not $\|Z_{k,l}\bar{y}\|_2^2$. Using $\|Z_{k,l}\bar{y}\|_2^2$ as a regularization term would in fact be equivalent to solving a different problem, namely: Tikhonov problem (1.2) with the identity as a regularization matrix (i.e., in standard form), in the preconditioned Krylov subspace $\mathcal{R}(Z_{k,l})$. It is important to note that $\mathcal{R}(Z_{k,l})$ incorporates regularization through preconditioning.

Flexible Krylov methods provide a natural framework to efficiently avoid nested loops of iterations by regarding the inverse of the regularization matrix (stemming from an iteratively reweighted regularization term) as iteration-dependent right preconditioning in (2.6). In this setting, at the kth iteration, the weights W_k are updated using the most recent approximation of the solution, i.e., the one at the $(k-1)$th iteration of the flexible solver, and incorporated in the construction of the flexible Krylov space in the form of the adaptive preconditioner W_k^{-1}. Flexible Krylov subspaces based on either the flexible Arnoldi or the flexible Golub-Kahan decompositions are summarized below.

Flexible Arnoldi decomposition. The flexible Arnoldi decomposition of $A \in \mathbb{R}^{n\times n}$ was first introduced in [40], and it is commonly employed in different settings to incorporate adaptive or increasingly improved preconditioners into the solution subspace; see [42, Chapter 9] and [43, 44]. Given A (square), b and right iteration-dependent preconditioning matrices W_k^{-1}, the partial factorization

$$(2.11) \qquad AZ_k = V_{k+1}\bar{H}_k,$$

is updated at iteration k (for $k \leq n$), where $\bar{H}_k \in \mathbb{R}^{(k+1)\times k}$ is upper Hessenberg, V_{k+1} has orthonormal columns with $v_1 = b/\|b\|_2$, and $Z_k = [W_1^{-1}v_1, ..., W_k^{-1}v_k] \in \mathbb{R}^{n\times k}$. Note that, when the preconditioning is fixed, i.e., $W_i = W$, flexible Arnoldi reduces to standard right-preconditioned Arnoldi (see equation (2.9)).

Flexible Golub-Kahan decomposition. The flexible Golub-Kahan decomposition of $A \in \mathbb{R}^{m\times n}$ has been recently introduced in [9] to solve ℓ_p-regularized least squares problems. Given A, b, and iteration dependent right preconditioning matrices $(W_k^{-1})^2$, the partial factorizations

$$(2.12) \qquad AZ_k = U_{k+1}M_k \quad \text{and} \quad A^TU_{k+1} = V_{k+1}S_{k+1}$$

are updated at iteration k (for $k \leq \min\{m, n\}$). In the first equation of (2.12), $M_k \in \mathbb{R}^{(k+1)\times k}$ is upper Hessenberg, $U_{k+1} \in \mathbb{R}^{m\times(k+1)}$ has orthonormal columns with $u_1 = b/\|b\|_2$, and $Z_k = [(W_1^{-1})^2v_1, ..., (W_k^{-1})^2v_k] \in \mathbb{R}^{n\times k}$. Moreover, $S_{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$ is upper triangular and $V_{k+1} \in \mathbb{R}^{n\times(k+1)}$ has orthonormal columns. Note that, for fixed preconditioning, i.e., $W_i = W_k$, FLSQR with preconditioner $(W_k^{-1})^2$ reduces to right preconditioned LSQR, which is mathematically equivalent to CG applied to the normal equations with split preconditioner W_k^{-1}. Although this relation is not stressed in [9], it can be observed in the definition of the search space for preconditioned LSQR in equation (2.10). The cost of computing these partial factorizations is dominated by one matrix vector product with A and one matrix vector product with A^T per iteration.

Detailed computations to update the partial flexible Arnoldi and flexible Golub-Kahan decompositions at the kth iteration are reported below. Notation-wise, $[\cdot]_{i,j}$ denotes the (i,j)th entry of the a matrix, and the vectors v_i, u_i, and z_i denote the ith column of the matrices V_k, U_k, and Z_k, correspondingly.

Flexible Arnoldi update

$z_k = W_k^{-1} v_k$
$w = A z_k$
Compute $[H]_{i,k} = w^T v_i$ for $i = 1, \dots, k$ and set $w = w - \sum_{i=1}^{k} [H]_{i,k} v_i$
Set $[H]_{k+1,k} = ||w||_2$ and, if $[H]_{k+1,k} \neq 0$, take $v_{k+1} = w/[H]_{k+1,k}$

Flexible Golub-Kahan update

$w = A^T u_k$
Compute $[S]_{i,k} = w^T v_i$ for $i = 1, \dots, k-1$ and set $w = w - \sum_{i=1}^{k-1} [S]_{i,k} v_i$
Set $[S]_{k,k} = ||w||_2$ and, if $[S]_{k,k} \neq 0$, take $v_k = w/[S]_{k,k}$
$z_k = (W_k^{-1})^2 v_k$
$w = A z_k$
Compute $[M]_{i,k} = w^T u_i$ for $i = 1, \dots, k$ and set $w = w - \sum_{i=1}^{k} [M]_{i,k} u_i$
Set $[M]_{k+1,k} = ||w||_2$ and, if $[M]_{k+1,k} \neq 0$, take $u_{k+1} = w/[M]_{k+1,k}$

Flexible methods to solve ℓ_p-regularized least square problems have already been used in [18, 9], where, at the kth iteration, the following projected problem is solved:

$$\bar{y}_k = \arg\min_{\bar{y}} \|A Z_k \bar{y} - b\|_2^2 + \lambda \|\bar{y}\|_2^2 , \quad \text{so that} \quad x_k = Z_k \bar{y}_k \, . \tag{2.13}$$

Note that $\bar{y}_k$ corresponds to the coefficients of the solution of (1.2) (in standard form) in the basis given by the columns of Z_k, which span a flexible Krylov space of dimension k with iteration dependent preconditioner W_k^{-1} and $(W_k^{-1})^2$ for FGMRES and FLSQR, respectively, where $W_k = \widetilde{W}^{(p,\tau)}(x_{k-1})$. Although extensive numerical tests show that methods (2.13) are efficient and deliver excellent reconstructions when compared to other Krylov solvers and other state-of-the-art methods for (1.5), it should be noted that solving problem (2.13) is not equivalent to solving problem (2.5) projected onto an appropriate flexible Krylov subspace at the kth iteration. Indeed, assume that n iterations of a flexible algorithm (2.13) have been performed, so that $\mathcal{R}(Z_n) = \mathbb{R}^n$: in this situation expression (2.13) corresponds to the Tikhonov problem (1.2) in standard form associated to (1.1) (and not the modification of the ℓ_2-ℓ_p problem in (2.4)). In other words, the "first-regularize-then-project" approach is not equivalent to the "first-project-then-regularize" approach for flexible Krylov solvers. Alternatively, this mismatch can be explained using the fact that, unlike in the case of (non flexible) preconditioned Krylov methods, in the problem projected using flexible Krylov subspaces there is no straightforward way of representing the variable $\bar{x}$ in (2.6) before "back-transformation". Note that [9] proposes to replace the regularization term $\|\bar{y}\|_2^2$ in (2.13) by $\|Z_k \bar{y}\|_2^2$: while (2.13) can be regarded as a hybrid regularization method that imposes additional standard form Tikhonov regularization on the projected solution $\bar{y}_k$, the regularization term $\|Z_k \bar{y}\|_2^2$ enforces standard form

Tikhonov regularization on $x_k = Z_k\bar{y}_k$ and does not lead to a scheme equivalent to the "first-regularize-then-project" one, either.

In the following section, two algorithms exploiting flexible Krylov subspaces in connection with the "first-regularize-then-project" framework will be presented along with a proof of convergence of the resulting schemes.

3. Iteratively Reweighted Flexible Krylov Subspace Methods.

In this section, two new algorithms are presented to solve (2.4) using a sequence of approximate problems of the form (2.5) and flexible Krylov subspaces (based on the flexible Arnoldi decomposition and the flexible Golub-Kahan decomposition respectively).

Here and in the following, without loss of generality, no initial guess is considered for the solution of (2.4) in a "warm start" fashion; however, a possible initial guess $x_0 \neq 0$ may be purely used to initialize the weights (2.3) at the very first iteration of the algorithm. The presented algorithms are assumed to be breakdown-free, i.e., at iteration $k \leq \min\{m, n\}$, the approximation subspace $\mathcal{R}(Z_k)$ for the solution has dimension k.

3.1. The new IRW-FGMRES and IRW-FLSQR methods.

The kth iteration of the new IRW-FGMRES or IRW-FLSQR methods computes an approximate solution x_k belonging to the space spanned by the columns of the matrix Z_k appearing in (2.11) or (2.12), respectively. More precisely, problem (2.5) is solved partially (i.e., in the space spanned by the columns of Z_k) as a projected least squares problem of the form

$$\bar{y}_k = \arg\min_{\bar{y}} \|AZ_k\bar{y} - b\|_2^2 + \lambda\|W_kZ_k\bar{y}\|_2^2\,, \quad \text{so that} \quad x_k = Z_k\bar{y}_k\,. \tag{3.1}$$

Let

$$W_kZ_k = Q_kR_k, \quad \text{with} \quad Q_k \in \mathbb{R}^{n\times k}, \quad R_k \in \mathbb{R}^{k\times k} \tag{3.2}$$

be the reduced QR factorization of the tall and skinny matrix W_kZ_k, which can be computed efficiently (see, for example, [13]). Then (3.1) is equivalent to

$$\bar{y}_k = \arg\min_{\bar{y}} \|\bar{H}_k\bar{y} - \|b\|_2e_1\|_2^2 + \lambda\|R_k\bar{y}\|_2^2\,, \quad \text{so that} \quad x_k = Z_k\bar{y}_k, \tag{3.3}$$

for IRW-GMRES, or

$$\bar{y}_k = \arg\min_{\bar{y}} \|M_k\bar{y} - \|b\|_2e_1\|_2^2 + \lambda\|R_k\bar{y}\|_2^2\,, \quad \text{so that} \quad x_k = Z_k\bar{y}_k, \tag{3.4}$$

for IRW-FLSQR. With a notation analogous to equation (2.13), $\bar{y}_k$ corresponds to the coefficients of the solution of (2.5) in the basis formed by the columns of Z_k, which span a flexible Krylov space of dimension k with iteration dependent preconditioning W_k^{-1} for IRW-FGMRES and $(W_k^{-1})^2$ for IRW-FLSQR (where $W_k = \widetilde{W}^{(p,\tau)}(x_{k-1})$). After the approximate solution x_k to problem (3.1) has been computed, the weights $W_{k+1} = \widetilde{W}^{(p,\tau)}(x_k)$ are (immediately) updated to be used in the next IRW-FGMRES or IRW-FLSQR iteration.

Although (3.1) might seem a rather unnecessarily convoluted formulation, since a change of variables for the regularization term is done and undone (i.e., an initial transformation into standard form in (2.6) eventually leads to a Tikhonov problem in general form), formulation (3.1) provides two main advantages over (2.8) and other

IRN strategies based on Krylov subspaces. Firstly, the iteration dependent regularization matrix W_k favorably affects the approximation subspace for the solution of problems of the form (2.5), i.e.,

$$x_k \in \mathcal{R}(Z_k) = \mathcal{R}([W_1^{-1}v_1, ..., W_k^{-1}v_k]),$$

for a set of vectors v_i that depend on the choice of IRW-FGMRES or IRW-FLSQR; see also [9, 20]. Secondly, problem (3.1) can be interpreted as a projection of the kth full-dimensional Tikhonov problem (2.5) (i.e., in a "first-regularize-then-project" framework). As a consequence, it can be proven that the sequence of approximate solutions $\{x_k\}_{k\geq 1}$ computed by IRW-FGMRES or IRW-FLSQR converges to the solution of problem (2.4).

Remark 3.1. Note that, assuming $n \leq m$ in (1.1), the IRW-FGMRES and IRW-FLSQR methods can be extended to the case when the number of iterations exceeds n by considering

$$x_k = \begin{cases} \arg\min_{x\in\mathcal{R}(Z_k)} T_k(x), & \text{for} \quad k = 1, ..., n-1 \\ \arg\min_{x\in\mathbb{R}^n} T_k(x), & \text{for} \quad k = n, ... \end{cases} \tag{3.5}$$

where $T_k(x)$ is defined in (2.5). Indeed, when $n \leq k$, an iteration of IRW-FGMRES or IRW-FLSQR corresponds to an IRN iteration for ℓ_p regularization (1.5), where the solution of each subproblem (2.5) is computed in a 'direct' fashion because the approximation subspace for the solution coincides with $\mathbb{R}^n$. Note however that this situation is not expected to happen in practice for large-scale problems.

Remark 3.2. Some numerical instabilities might happen in generating $W_k Z_k$ in the regularization term in (3.1) when applying the new IRW-FGMRES and IRW-FLSQR methods, due to division by almost zeros in the weights component. Section 4 presents an example where this happens, and discusses two possible fixes that can be adopted at implementation level to improve stability.

The new IRW-FGMRES and IRW-FLSQR methods are sketched in Algorithm 3.1.

Algorithm 3.1 IRW-FGMRES and IRW-LSQR methods.

Input: A, b, p, $\tau > 0$, x_0
Initialize: $v_1 = b/||b||_2$ for IRW-FGMRES, $u_1 = b/||b||_2$ for IRW-FLSQR
If $x_0 \neq 0$ $W_1 = \widetilde{W}^{(p,\tau)}(x_0)$ else $W_1 = \mathrm{I}_n$
for $k = 1, \ldots,$ until a stopping criterion is satisfied **do**
  Update (2.11) (for IRW-FGMRES) or (2.12) (for IRW-FLSQR)
  Compute $\bar{y}_k$ in (3.3) (for IRW-FGMRES) or in (3.4) (for IRW-FLSQR)
  Compute $x_k = Z_k \bar{y}_k$
  Update the weights $W_{k+1} = \widetilde{W}^{(p,\tau)}(x_k)$
end for

If $k \ll \min\{m, n\}$, the computational cost of the kth iteration of Algorithm 3.1 is dominated by the computational cost of updating the factorizations (2.11) or (2.12). Indeed, for IRW-FGMRES and assuming that A is dense, computing matrix-vector products with A amounts to $O(mn)$ flops (but could be much less if A is sparse or has some structure), while performing the orthonormalization steps amounts to $O(kn)$ flops. Forming the matrix $W_k Z_k$ and computing the QR factorization (3.2) amounts to $O(nk^2)$ flops, while solving problem (3.3) and forming x_k amounts to $O(k^3)$ flops. Similar estimates can be derived for IRW-FLSQR.

3.2. Convergence of IRW-FGMRES and IRW-FLSQR. Note that, even if in practice IRW-FGMRES and IRW-FLSQR allow for an iteration-dependent choice of the regularization parameter λ in the functional $T^{(p,\tau)}(x)$ in (2.4), in this section λ is assumed to be known a priori and fixed throughout the iterations.

LEMMA 3.3. *Assume that no breakdown happens in the flexible Arnoldi and Golub-Kahan algorithms. Then the sequence* $\{T^{(p,\tau)}(x_k)\}_{k\geq 1}$ *for* $0 < p \leq 2$*, where* $T^{(p,\tau)}(x)$ *is defined in (2.4), and where* x_k *is the approximate solution computed after* k *steps of the IRW-FGMRES or the IRW-FLSQR methods, is decreasing monotonically and it is bounded from below by zero.*

Proof. Consider a fixed $p \in (0, 2]$ and $\tau > 0$. Since $T^{(p,\tau)}(x) \geq 0$, only the fact that $T^{(p,\tau)}(x_k)$ is monotonically decreasing needs to be proved, i.e., that $T^{(p,\tau)}(x_k) \leq T^{(p,\tau)}(x_{k-1})$ for every $k \geq 1$. Consider $T_k(x)$ defined in (2.5) (note that it is defined with respect to $W_k = \widetilde{W}^{(p,\tau)}(x_{k-1})$) and recall that $T_k(x)$ is a quadratic tangent majorant of $T^{(p,\tau)}(x)$ at point x_{k-1}, i.e.,

$$T^{(p,\tau)}(x_{k-1}) = T_k(x_{k-1}) \quad \text{and} \quad T^{(p,\tau)}(x) \leq T_k(x) \quad \forall x. \tag{3.6}$$

In particular, for x_k,

$$T^{(p,\tau)}(x_k) \leq T_k(x_k). \tag{3.7}$$

Moreover, recalling the definition of x_k in (3.1), and since $x_{k-1} \in \mathcal{R}(Z_{k-1}) \subset \mathcal{R}(Z_k)$,

$$T_k(x_k) = \min_{x \in \mathcal{R}(Z_k)} T_k(x) \leq T_k(x_{k-1}), \tag{3.8}$$

so, combining equations (3.6), (3.7) and (3.8),

$$T^{(p,\tau)}(x_k) \leq T_k(x_k) \leq T_k(x_{k-1}) = T^{(p,\tau)}(x_{k-1})\,, \tag{3.9}$$

which concludes the proof. ∎

THEOREM 3.4. *Under the same assumptions of Lemma 3.3, the sequence* $\{x_k\}_{k\geq 1}$*, where* x_k *is the approximated solution computed after* k *steps of IRW-FGMRES or IRW-FLSQR with* $p > 0$*, is such that*

$$\lim_{k\to\infty} \|x_k - x_{k-1}\|_2 = 0.$$

Moreover, it converges to a stationary point of $T^{(p,\tau)}$ *and, if* $p \geq 1$*, this is the unique solution of (2.4).*

Proof. Thanks to Lemma 3.3, $\{T^{(p,\tau)}(x_k)\}_{k\geq 1}$ has a stationary point. The convergence result for $\{x_k\}_{k\geq 1}$ proved in Theorem 5 of [27] for majorization-minimization methods based on Generalized Krylov subspaces, when $k \geq n$, can be applied in this setting as the same majorization for $T^{(p,\tau)}$ is used. ∎

It should be stressed that, although the regularization parameter λ in (3.1) is assumed fixed, the IRW-FGMRES and the IRW-FLSQR methods naturally allow for an iteration-dependent regularization parameter λ_k to be adaptively set at the kth iteration (e.g., at line 6 of Algorithm 3.1). Indeed, when considering inner-outer iterative schemes for (2.6) or flexible Krylov methods for (2.13), one can employ approaches typically used for hybrid methods (e.g., projected versions or approximations of well-known regularization parameter rules for Tikhonov problem (1.2); see [9, 18]). For IRW-FGMRES and IRW-FLSQR to be consistent with the "first-regularize-then-project" framework, one should make sure that the parameter λ_k selected at the kth iteration according to the adopted rule is a suitable λ for problem (2.5) and, eventually, for problem (1.5): although for projection methods based on standard Krylov subspaces convergence of λ_k to a λ can be guaranteed in some situations (e.g., when using standard Golub-Kahan bidiagonalization and the discrepancy principle, see [21]), it is not immediate to generalize these results to IRW-FGMRES and IRW-FLSQR. In the numerical experiments displayed in Section 4 the discrepancy principle is employed to select the regularization parameter at each IRW-FGMRES or IRW-FLSQR iteration.

3.3. Alternative interpretation of IRW flexible methods. Augmented Krylov subspaces are most commonly used to incorporate an initial 'guess' subspace of moderate dimension within a (traditional) Krylov subspace for the approximation of the solution of a linear system. In the framework of ill-posed problems, this approach is extremely beneficial if the initial 'guess' vectors are chosen to model known features of the solution (see, e.g, [1, 2, 3, 15]); a combination of Tikhonov regularization and projection onto augmented Krylov subspaces has been considered in [24]. When performing iteratively reweighted schemes, a sequence of different but closely related problems of the form (2.5) or, equivalently, (2.6), is considered. Potentially, an augmented Krylov subspace method could be used to solve each of the problems if one had a good initial set of 'guess' vectors. In this setting it is argued that IRW flexible Krylov methods can be regarded as particular instances of augmented Krylov methods where, when approximating the solution of the kth problem of the form (2.5) (i.e., at iteration $k \leq \min\{m, n\}$), the initial 'guess' subspace is taken to be $\mathcal{R}(Z_{k-1})$ (i.e, the flexible Krylov subspace available from the previous iteration) and only one iteration of a (standard) Krylov method is performed (so that, in particular, the size of the augmentation subspace for the kth problem of the form (2.5) is $k-1$). This interpretation also draws similarities with the idea of recycling Krylov methods for sequences of linear systems [29, 38], and can be extended to flexible Krylov methods in general. Indeed, some analogies between flexible GMRES and augmented GMRES were already established in [8, 41]. Although the following derivations are specified for IRW-FGMRES and for augmented methods based on GMRES, they can be easily extended to handle IRW-FLSQR and augmented methods based on LSQR.

Consider the kth IRW-FGMRES iteration. Using the identity

$$Z_k = [Z_{k-1}, W_k^{-1} v_k] = W_k^{-1}[W_k Z_{k-1}, v_k]\,,$$

the flexible Arnoldi partial factorization (2.11) can be reformulated as

$$A[Z_{k-1}, W_k^{-1} v_k] = A W_k^{-1}[W_k Z_{k-1}, v_k] = [V_k, v_{k+1}]\bar{H}_k, \tag{3.10}$$

and the kth minimization problem (3.1), solved at the kth iteration of IRW-FGMRES, can be expressed as

$$\bar{y}_k = \arg\min_{\bar{y}} \|A W_k^{-1}[W_k Z_{k-1}, v_k]\bar{y} - b\|_2^2 + \lambda \|[W_k Z_{k-1}, v_k]\bar{y}\|_2^2\,. \tag{3.11}$$

Then, $\bar{x}_k = [W_k Z_{k-1}, v_k]\bar{y}_k$ is an approximate solution of the kth problem of the form (2.6) that belongs to the space $\mathcal{R}([W_k Z_{k-1}, v_k])$, and $x_k = W_k^{-1}\bar{x}_k$ is an approximate solution of the kth problem of the form (2.5) that belongs to the space $\mathcal{R}(Z_k)$.

Now consider a single step of the augmented Arnoldi process with augmentation space $\mathcal{R}(Z_{k-1})$ and with starting vector

$$\hat{v}_k = (I - V_{k-1}V_{k-1}^T)r_{k-1}/\|(I - V_{k-1}V_{k-1}^T)r_{k-1}\|_2, \quad \text{with} \quad r_{k-1} = b - Ax_{k-1}\,, \tag{3.12}$$

so that $\hat{v}_k = v_k$. This leads to an approximation subspace for the solution of dimension k, and can be written as

1: Define $\hat{v}_k$ as in (3.12) and set $V_k = [V_{k-1},\, \hat{v}_k]$.
2: Compute $\hat{z}_k = W_k^{-1}\hat{v}_k$.
3: Compute $\hat{w} = (I - V_k V_k^T)A\hat{z}_k$.
4: Take $[\hat{H}]_{k+1,k} = \|\hat{w}\|_2$.
5: Compute $\hat{v}_{k+1} = \hat{w}/[\hat{H}]_{k+1,k}$.

In the above algorithm, the matrix V_k in line 1 coincides with the matrix V_k in (3.10) because $\hat{v}_k = v_k$. Lines 3 to 5 can be rearranged as

$$[\hat{H}]_{k+1,k}\,\hat{v}_{k+1} = (I - V_k V_k^T)A\hat{z}_k\,, \quad \text{so that} \quad A\hat{z}_k = V_k(V_k^T A\hat{z}_k) + \hat{v}_{k+1}[\hat{H}]_{k+1,k}\,.$$

Incorporating augmentation and considering the partial factorization (2.11) with k replaced by $k-1$, the following decomposition is obtained

$$A[Z_{k-1}, \hat{z}_k] = [V_k, \hat{v}_{k+1}] \begin{bmatrix} \bar{H}_{k-1} & V_k^T A\hat{z}_k \\ 0 & [\hat{H}]_{k+1,k} \end{bmatrix} = [V_k, \hat{v}_{k+1}]\hat{H}_k. \tag{3.13}$$

Comparing the above algorithm to the flexible Arnoldi algorithm in Section 2, it is immediate to see that $\hat{z}_k = W_k^{-1}\hat{v}_k = W_k^{-1}v_k = z_k$, and $\hat{v}_{k+1} = v_{k+1}$. Therefore, by inspection, it can be seen that this formulation is equivalent to (3.10), and that $\bar{H}_k = \hat{H}_k$.

As a consequence, the projection step performed to compute $\bar{y}_k$ in (3.11) using either the flexible or the augmented approaches is equivalent, so the same kth approximate solution x_k of (3.1) is obtained.

The augmented method (3.13) mainly differs from the available augmented methods in the starting vector that is chosen for building the (standard) Krylov subspace: indeed, the latter either take the normalized right hand side b (i.e., the (standard) Krylov subspace is built first, and then enriched with the initial 'guess' subspace; see [15, 24]) or the orthogonal projection of b on the orthogonal complement of the initial 'guess' subspace (i.e., the (standard) Krylov subspace is built preserving orthogonality to the initial 'guess' subspace; see [1, 2, 3]). Note that the choice of the initial vector (3.12) for IRW-FGMRES more radically stems from the fact that $(I - V_k V_k^T)b = 0$, as $b \in \mathcal{R}(V_k)$.

The decomposition (3.13) associated to IRW-FGMRES is also analogous to the decompositions typically associated to recycling methods [38], the only difference being in the way the solution is computed (recycling often considers 'warm restarts', where computing the solution at the kth iteration amounts to computing the correction of an initial guess).

4. Numerical Experiments.

In this section the results of three experiments concerned with imaging problems are presented to illustrate the behaviour of the new methods. In all the experiments, x is the vector obtained by stacking the columns of a two dimensional discrete image. The new IRW-FGMRES and IRW-FLSQR methods are compared with other state-of-the-art solvers for (1.5) with $0 < p \leq 2$, including: other solvers based on generalized and flexible Krylov methods, first-order optimization methods or optimization methods based on quadratic separable approximations of part of the objective function, solvers that employ standard or preconditioned Krylov methods based on the Arnoldi and the Golub-Kahan bidiagonalization algorithms. To the best of our knowledge, comparisons between methods based on flexible and generalized Krylov subspaces have never been considered before. Table 1 summarizes the methods considered in this section, providing acronyms and brief descriptions thereof. Note that, for all the considered examples, the computation of matrix-vector products with A and, possibly, A^T dominates the computational cost of each iteration of all the methods listed in Table 1. In particular, Krylov methods based on the (flexible) Golub-Kahan algorithm (i.e., IRW-FLSQR, IRN-hLSQR, (hybrid) FLSQR) have the same computational cost per iteration as GKSpq, FISTA, and SpaRSA, since they require one matrix-vector product with A and A^T; Krylov methods based on the (flexible) Arnoldi algorithm (i.e., IRW-FGMRES, IRN-hGMRES, (hybrid) FGMRES) are the ones with the lowest cost per iteration, since they require only one matrix-vector product with A. As a consequence, in the following tests, methods that require fewer iterations to compute solutions of comparable qualities have to be regarded as more efficient.

Table 1: Summary of the methods considered in this section for approximating the solution of problem (1.5).

Method	Description	Note	References	Marker
IRW-FGMRES **IRW-FLSQR**	the new Algorithm 3.1	adaptive reg. parameter selection	–	blue line
IRN-hGMRES **IRN-hLSQR**	IRN strategy within an inner-outer scheme	preconditioned hybrid GMRES or LSQR is used to solve (2.6) at each outer iteration; adaptive reg. parameter selection	[39]	green line
hybrid FGMRES **hybrid FLSQR**	hybrid versions of FGMRES or FLSQR	standard form Tikhonov regularization applied on the projected solution; adaptive reg. parameter selection	[9, 18]	pink line
FGMRES **FLSQR**	Flexible GMRES or LSQR with sparsity-enforcing iteration-dependent preconditioning	no Tikhonov regularization for the projected problem	[9, 18]	dark red line
GKSpq	Generalized Krylov Subspace methods	initial subspace $\mathcal{K}_l(A^TA, A^Tb)$ with $l = 5$; adaptive reg. parameter selection	[31]	light blue line
FISTA	Fast ISTA	accelerated first-order optimization method	[4]	purple line
SpaRSA	Sparse Reconstruction by Separable Approximation	quadratic separable approximations of part of the objective function	[46]	orange line

When a method allows the regularization parameter λ to be adaptively set at each iteration, this is done according to the discrepancy principle [34] as described below. Assuming that a good approximation of the 2-norm of the noise vector e appearing in (1.1) is available, a zero-finder is employed to solve the following nonlinear equation with respect to $\lambda \geq 0$ at the kth iteration

$$\|Ax_k(\lambda) - b\|_2 = \eta\|e\|_2\,, \tag{4.1}$$

where $x_k(\lambda)$ is the approximate solution at iteration k given as a function of the regularization parameter λ, and $\eta \geq 1$ is a safety parameter. Note that equation (4.1) is guaranteed to have a solution as soon as $\|Ax_k(0)-b\|_2 \leq \|e\|_2$. For IRW-FGMRES,

$$\begin{aligned} x_k(\lambda) = Z_k\bar{y}_k &= Z_k(\bar{H}_k^T\bar{H}_k + \lambda R_k^T R_k)^{-1}\bar{H}_k^T\|b\|_2 e_1 \\ &= Z_k(\bar{H}_k^T\bar{H}_k + \lambda R_k^T R_k)^{-1}\bar{H}_k^T V_{k+1}^T b, \end{aligned} \tag{4.2}$$

where $\bar{H}_k$ is defined in equation (2.11) and R_k is obtained computing the reduced QR factorization of $W_k Z_k$; see (3.2). Then

$$\begin{aligned} \|Ax_k(\lambda) - b\|_2 &= \|AZ_k(\bar{H}_k^T\bar{H}_k + \lambda R_k R_k^T)^{-1}\bar{H}_k^T V_{k+1}^T b - b\|_2 \\ &= \|V_{k+1}\bar{H}_k(\bar{H}_k^T\bar{H}_k + \lambda R_k R_k^T)^{-1}\bar{H}_k^T V_{k+1}^T b - b\|_2 \\ &= \|\bar{H}_k(\bar{H}_k^T\bar{H}_k + \lambda R_k R_k^T)^{-1}\bar{H}_k^T\|b\|_2 e_1 - \|b\|_2 e_1\|_2, \end{aligned} \tag{4.3}$$

so that applying the discrepancy principle (4.1) does not require performing any additional matrix-vector product with A per iteration. An analogous argument can be made specifically for IRW-FLSQR (as expression (4.3) formally holds for IRW-FLSQR after replacing the matrix $\bar{H}_k$ by M_k), as well as for most of the algorithms listed above; see also [30, 19]. Note that, although synthetic noise e with known $\|e\|_2$ is always used in the following, estimates of the noise level or alternative parameter choice strategies that do not require an estimate of $\|e\|_2$ can be used if $\|e\|_2$ is not immediately available; see, e.g., [21, 45]. When no adaptive regularization parameter choice is supported (e.g., for FISTA and SpaRSA), the value of the regularization parameter computed by IRW-FGMRES or IRW-FLSQR (upon iteration termination) is used. Alternatively, such solvers can be run from scratch for different preselected values of the regularization parameter and the best solution can be picked according to some criterion, resulting in a very computationally demanding strategy.

Throughout all the experiments, if not stated otherwise, the values $p = 1$ and $\tau = 10^{-10}$ are chosen in (2.3), $\eta = 1$ is chosen in (4.1), and all the solvers are set to perform 200 (total) iterations. Although, provided that a suitable value of the regularization parameter is set at each iteration, the quality of the reconstructions computed by the new methods does not significantly deteriorate as the iterations proceed, one or more stopping criteria should be set in practice. A reasonable choice is to stop at the first iteration k such that

$$\frac{|\lambda_k - \lambda_{k-1}|}{\lambda_k} < \theta_1 \quad \text{or} \quad \frac{|s(x_k) - s(x_{k-1})|}{s(x_k)} < \theta_2 \tag{4.4}$$

where $\theta_1, \theta_2 > 0$ are user-selected thresholds, and where $s(\cdot)$ is a (practical) measure of the sparsity of the solution. In the following, given a vector y,

$$s(y) = \#\left\{i \;:\; |[y]_i| \geq 10^{-3}||y||_2\right\}\,, \quad \text{where } \# \text{ denotes cardinality.} \tag{4.5}$$

Stopping criteria (4.4) monitor the stabilization of some relevant quantities for the solution, so that one can expect x_k not to vary too much once they are satisfied; see [19]. In all the graphs presented below, the iteration satisfying the first stopping criterion in (4.4) with $\theta_1 = 10^{-4}$ is marked by a circle, and the iteration satisfying the second stopping criterion in (4.4) with $\theta_2 = 10^{-10}$ is marked by a triangle.

Experiment 1. The first experiment is concerned with image deblurring. The `star_cluster` test problem from *Restore Tools* [35] is used to generate an exact test image of size 256×256 pixels (so $n = 65536$ in (1.1)) and a square blurring matrix modelling *spatially variant blur* (we refer interested readers to [36] for a discussion of how the matrix A is represented, and how matrix-vector products can be done efficiently). The measurements are corrupted by Gaussian white noise e of level $\|e\|_2/\|b_{true}\|_2 = 10^{-2}$. The setting for this example can be observed in Figure 1. Note that $s(x_{true}) = 470$, i.e., only approximately 0.07% of the pixels can be regarded as different from zero in practice, according to definition (4.5). This example has been mimicked from [18]. Since A is square, the performance of IRW-FGMRES can be tested.

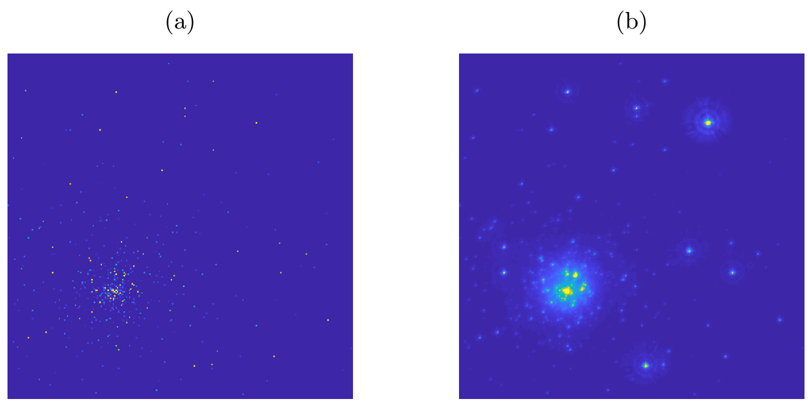

Fig. 1: *Experiment 1.* Setting for the `star_cluster` test problem. **(a)** True image x_{true}, **(b)** Noisy measurement b.

Figure 2 displays the behavior of the relative errors versus the number of iterations for the methods listed in Table 1. It can be observed in Figure 2 (a) that IRW-FGMRES shows a faster and more stable convergence when compared to other standard methods for ℓ_2-ℓ_p regularization. In particular, the new method stabilizes to roughly the same value of the relative error as IRN and FISTA, while SpaRSA converges to a reconstruction of worse quality. Even restricting the comparisons to other methods that build only one generalized or flexible Krylov subspace for the solution, the new IRW-FGMRES method shows a more desirable behavior. Indeed, it can be observed in Figure 2 (b) that the solver based on FGMRES displays some semiconvergence; this feature is shared by the hybrid version of FGMRES and may appear because a Tikhonov problem in standard form is solved, so that sparsity is only enforced through the construction of a suitable flexible Krylov subspace. Also, within the maximum number of allowed iterations, the quality of the solution computed by the solver based on generalized Krylov subspaces is lower than the IRW-FGMRES one: this shows that, for this test problem, the approximation subspace for the solution computed by IRW-FGMRES is better than the one computed by GKSpq.

Figure 3 (a) displays the values of the relative residuals $\|b - Ax_k(\lambda)\|_2/\|b\|_2$ versus the number of iterations k. One can clearly see that, since λ is adaptively set at each iteration using the discrepancy principle (for all the displayed methods except for FGMRES), the relative residual eventually stabilizes around the noise level, as it

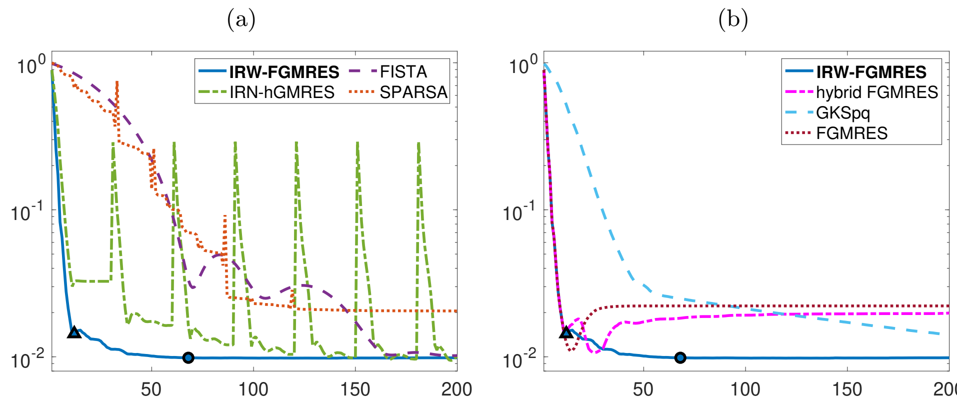

Fig. 2: *Experiment 1.* History of relative error norms (i.e., $\|x_k(\lambda) - x_{true}\|_2/\|x_{true}\|_2$ against iteration number k) for the new IRW-FGMRES, compared to **(a)** other standard solvers for the ℓ_2-ℓ_1 problem; **(b)** other flexible and generalized Krylov-based solvers. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

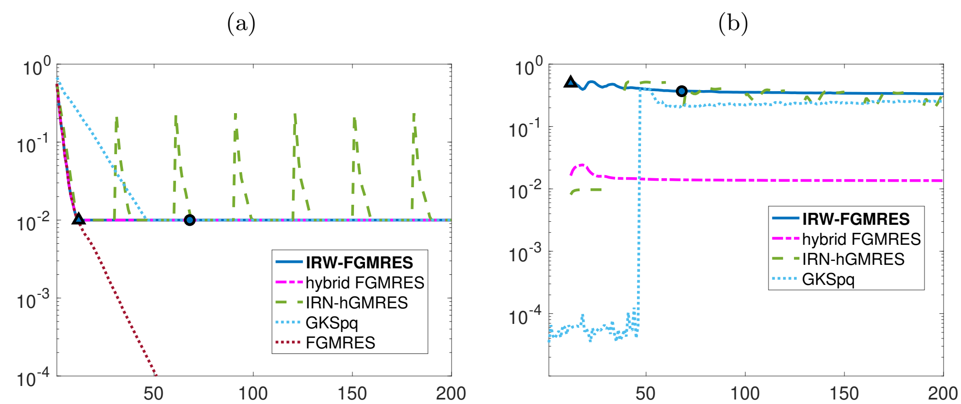

Fig. 3: *Experiment 1.* Methods based on Krylov subspaces. **(a)** History of the relative residuals. **(b)** History of the regularization parameters. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

should happen for regularization methods applied to ill-posed problems: this happens quite quickly for methods based on the flexible Arnoldi algorithm, but sensibly later for the GKSpq method (coherently to what is observed in Figure 2 (a)). Figure 3 (b) displays the values of the regularization parameters $\lambda = \lambda_k$ selected at each iteration versus the number of iterations k. It can be observed that the regularization parameter chosen by the new IRW-FGMRES method quickly stabilizes to a value that is similar to the one eventually selected by the IRN and the GKSpq methods. The regularization parameter chosen by the hybrid version of FGMRES stabilizes to

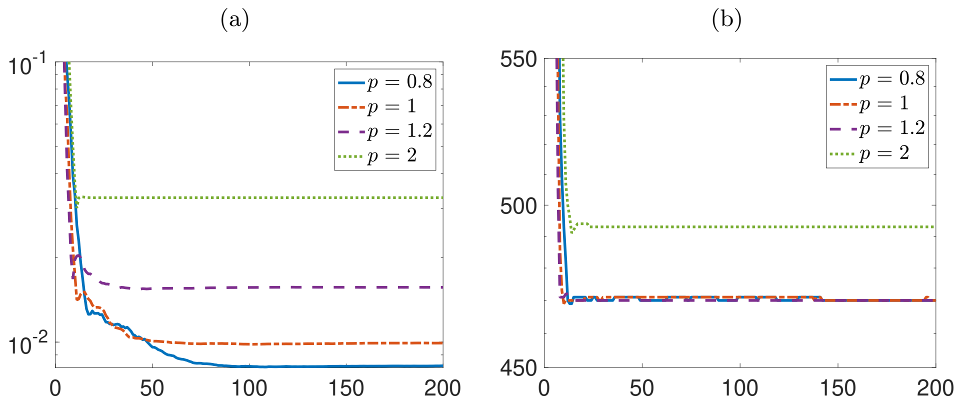

Fig. 4: *Experiment 1.* **(a)** History of the IRW-FGMRES relative error norms for different values of p in the ℓ_p regularization term. **(b)** History of $s(x_k)$ for IRW-FGMRES and for different values of p in the ℓ_p regularization term.

a different value, which is more similar to the one selected during the first IRN outer iteration, i.e., when a Tikhonov problem in standard form is solved. This behavior is consistent with the arguments presented in Sections 2 and 3. Indeed, similarly to IRN and GKSpq, IRW-FGMRES can be proved to converge to a stationary point of (2.4): therefore it should be expected that the regularization parameter adaptively selected by these methods according to the discrepancy principle also stabilizes around a common value. On the contrary, hybrid FGMRES imposes additional standard form Tikhonov regularization on the projected solution: therefore it should be expected that the regularization parameter stabilizes around a value suitable for standard form Tikhonov regularization.

Finally, Figure 4 (a) displays the history of relative errors obtained using IRW-FGMRES for different values of p in the ℓ_p regularization term. Note that, since the quality of the solution generally improves when taking $p < 1$ (coherently with the fact that x_{true} is very sparse), one can expect that IRN-FGMRES is converging to a global minimum when started with $x_0 = 0$ for this test problem. Correspondingly, Figure 4 (b) displays the values of $s(x_k)$ versus the number of iterations k. It can be observed that, when the value of p in the ℓ_p regularization term is 2, the recovered solution is considerably less sparse than x_{true}, whereas for smaller values of p, the value of $s(x_k)$ approximates $s(x_{true}) = 470$. In particular, note that, when $p = 1$, $s(x_k)$ converges to $s(x_{true}) = 470$ when using IRW-FGMRES. Even if not shown, this is also true for FISTA, SpaRSA, IRN-hGMRES, FGMRES, and hybrid FGMRES. Similarly, the solution obtained using the GKSpq method at the end of the iterations had a $s(x_k)$ of 472.

Experiment 2. The second test problem uses the so-called `hst` (Hubble space telescope) test image together with the spatially invariant `speckle medium blur` linear operator available within *IR Tools* [17]. The noise level is $\|e\|_2/\|b_{true}\|_2 = 10^{-2}$ and $\eta = 1$ is chosen in (4.1). The setting for this experiment can be observed in Figure 5. The object displayed in this test image is not as sparse as in the previous test problem; the overall sparsity is associated to the uniform (zero) background. Note that, in this example, the square matrix $A \in \mathbb{R}^{n\times n}$ (where $n = 65536$) is generated

603 by a highly anisotropic blur (see Figure 5 (b)): in this situation, there is no guarantee that GMRES can perform well; see [14]. For this reason, only the performance of methods based on LSQR will be compared.

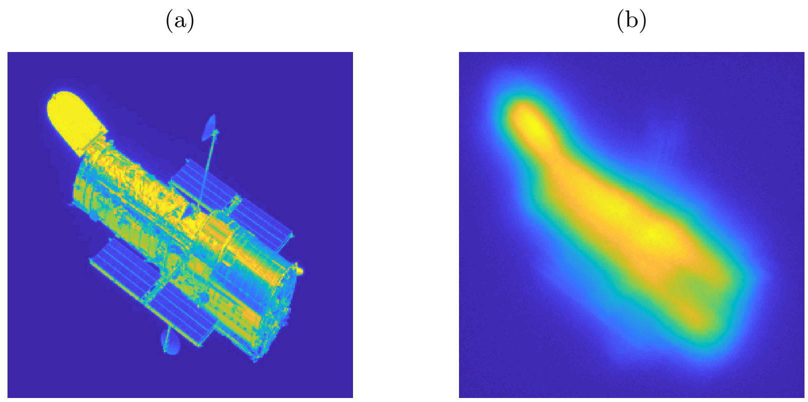

Fig. 5: *Experiment 2*. Setting for the `hst` test problem. **(a)** True image x_{true}, **(b)** Noisy measurement b.

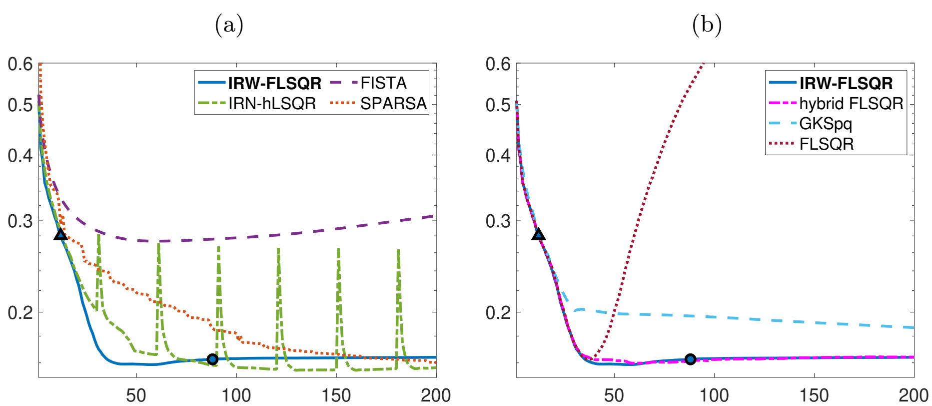

Fig. 6: *Experiment 2*. History of relative error norms for the new IRW-FLSQR, compared to **(a)** other standard solvers for the $\ell_2 - \ell_1$ problem; **(b)** other flexible and generalized Krylov-based solvers. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

606 The relative error history associated to different solvers for (2.4) is displayed in Figure 6. It should be stressed that, when running IRW-FLSQR for this experiment, $\tau = 0.01$ is set in (2.3) to avoid numerical instabilities happening in the generation of $W_k Z_k$ (as mentioned in Remark 3.2). As it can be seen in Figure 8 (a), a smaller value of τ would lead to solutions of worse quality. Alternatively, Figure 8 (b) shows the history of the relative errors when the components of the weights $W_k = \widetilde{W}^{(p,\tau)}(x_{k-1,\star})$ are set to 0 in (2.5) if they are higher than a certain threshold τ_W (as suggested in [39]).

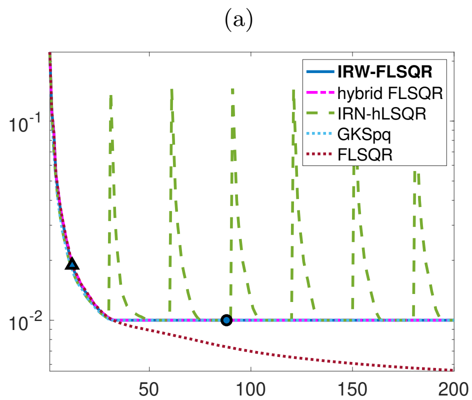

Fig. 7: *Experiment 2.* Methods based on Krylov subspaces. **(a)** History of the relative residuals. **(b)** History of the regularization parameters. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

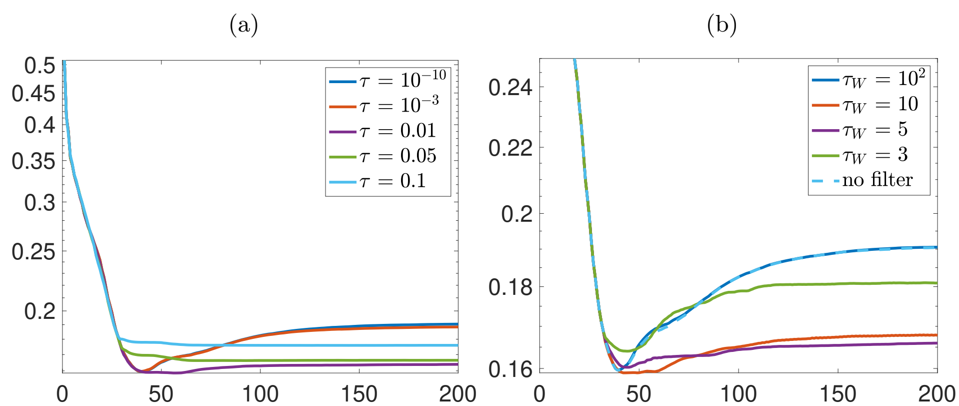

Fig. 8: *Experiment 2.* Different strategies to stabilize the quality of the solution. History of the relative error norms for the new IRW-FLSQR: **(a)** for different values of τ, **(b)** for different values of τ_W.

As in the previous example, Figure 7 (a) displays the values of the relative residuals $\|b - Ax_k(\lambda)\|_2/\|b\|_2$ versus the number of iterations k and Figure 7 (b) displays the values of the regularization parameters $\lambda = \lambda_k$ selected at each iteration k according to the discrepancy principle. The behavior of these quantities is very similar to the one observed in the previous example and it can be interpreted in the same way.

Experiment 3. This test problem models sparse X-ray tomographic reconstruction with oversampled data. The chosen test phantom is the `ppower` image from [26], generated in such a way that only 10% of its pixels are exactly non-zero; this phantom is also fairly smooth (see Figure 9 (a)). A measurement geometry consisting of 362 equidistant parallel beams rotated around 224 equidistant angles between 1° and

180° is considered. This corresponds to a discrete forward operator $A \in \mathbb{R}^{m\times n}$ with $m = 81088$ and $n = 65536$, so that only methods based on the Golub-Kahan decomposition can be compared. The noise level in this example is $\|e\|_2/\|b_{true}\|_2 = 1.5 \cdot 10^{-2}$.

(a) (b)

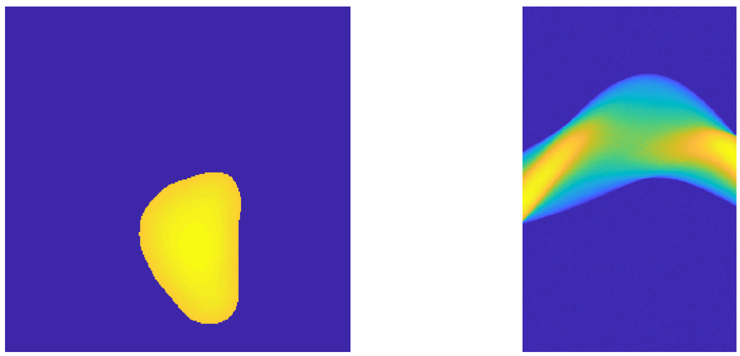

Fig. 9: *Experiment 3*. Setting for the `ppower` test problem. **(a)** True phantom x_{true}, **(b)** Noisy sinogram measurement b.

(a) (b)

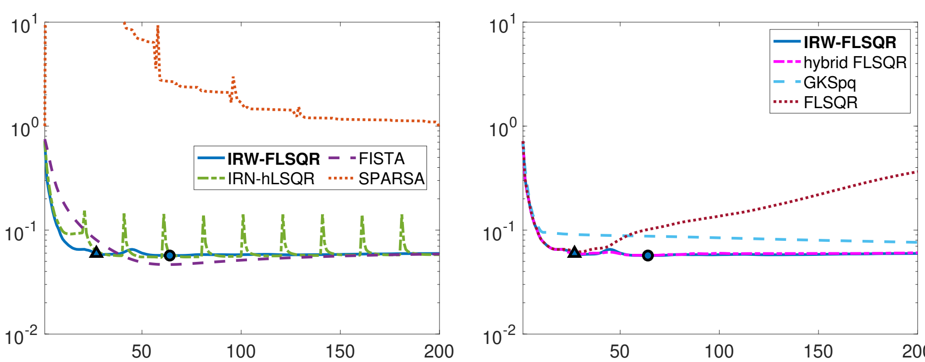

Fig. 10: *Experiment 3*. History of relative error norms for the new IRW-FLSQR, compared to **(a)** other standard solvers for the $\ell_2 - \ell_1$ problem; **(b)** other flexible and generalized Krylov-based solvers. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

The convergence results for this tomography example with oversampled data are displayed in Figures 10 and 11. The methods based on flexible Krylov subspaces all perform similarly well. FISTA seems to deliver a solution of slightly better quality than IRW-FLSQR, but it takes more iterations to do so. SpaRSA seems to perform poorly for this test problem; it may be expected that experimenting with different values of the regularization parameter could lead to an improved solution.

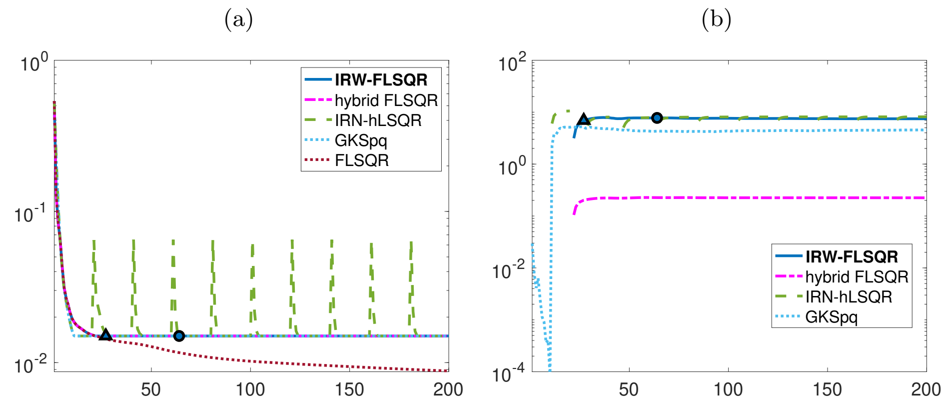

Fig. 11: *Experiment 3* Methods based on Krylov subspaces. **(a)** History of the relative residuals. **(b)** History of the regularization parameters chosen according to the discrepancy principle. The circle and triangle markers correspond to stopping criteria (4.4) based on the stabilization of λ and $s(x_k)$, respectively.

5. Conclusions. This paper presents two new algorithms, called IRW-FGMRES and IRW-FLSQR, that efficiently solve the ℓ_2-ℓ_p minimization problem (1.5) by partially solving a sequence of quadratic problems arising from the Iteratively Reweighted Norm (IRN) strategy. The new methods compute approximate solutions belonging to flexible Krylov subspaces of increasing dimension, that encode regularization through iteration-dependent "preconditioning", so to avoid nested loops of iterations and build only one approximation subspace for the solution. With respect to other available IRN solvers, the new approach not only improves the efficiency of the algorithm, but also avoids the need of choosing stopping criteria for the inner iterations. Moreover, the regularization parameter can be set adaptively along the iterations (even using strategies other than the discrepancy principle, which is considered in this paper). The new flexible Krylov solvers are supported by a solid theoretical justification: indeed, the sequence of approximate solutions given by Algorithm 3.1 is guaranteed to converge to the solution of the smoothed formulation (2.4) of problem (1.5).

Extensive numerical testing, involving large-scale inverse problems in imaging, shows that IRW-FGMRES and IRW-FLSQR are competitive with other standard implementations of IRN methods as well as other optimization methods. Moreover, although IRW-FGMRES can only be applied to a square coefficient matrix A and is not guaranteed to work well if A is highly non normal, it requires only a single matrix-vector product with A at each iteration, while IRW-FLSQR needs an additional matrix-vector product with A^T at each iteration. It is worth highlighting again that, although the hybrid implementations of FGMRES, FLSQR [18, 9] and IRW-FGMRES, IRW-FLSQR have a similar behavior in most of the performed numerical tests, the former still lack a solid theoretical justification of convergence.

Future work will include a theoretical investigation of the convergence of IRW-FGMRES and IRW-FLSQR in presence of a variable regularization parameter that is automatically set at each iteration according to a given rule, and the extension of the new IRW flexible Krylov methods to handle more involved regularizers, such as total variation and generalizations thereof.

REFERENCES

[1] J. Baglama and L. Reichel, *Augmented GMRES-type methods*, J. Numer. Linear Algebra Appl., 14 (2007), pp. 337 – 350, https://doi.org/10.1002/nla.518.

[2] J. Baglama and L. Reichel, *Decomposition methods for large linear discrete ill-posed problems*, J. Comput. Appl. Math., 198 (2007), pp. 332–343, https://doi.org/10.1016/j.cam.2005.09.025.

[3] J. Baglama, L. Reichel, and D. Richmond, *An augmented LSQR method*, Numer. Algorithms, 64 (2013), https://doi.org/10.1007/s11075-012-9665-8.

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, https://doi.org/10.1137/080716542.

[5] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484.

[6] A. Buccini, M. Donatelli, and L. Reichel, *Iterated Tikhonov regularization with a general penalty term*, Numer. Linear Algebra Appl., 24 (2017), p. e2089, https://doi.org/10.1002/nla.2089.

[7] D. Calvetti, *Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective.*, J. Comput. Appl. Math., 198 (2007), p. 378–395.

[8] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1996), https://doi.org/10.1002/(SICI)1099-1506(199701/02)4:1⟨43::AID-NLA99⟩3.0.CO;2-Z.

[9] J. Chung and S. Gazzola, *Flexible Krylov methods for ℓ_p regularization*, SIAM J. Sci. Comput., 41 (2019), pp. S149–S171, https://doi.org/10.1137/18M1194456.

[10] J. Chung, M. E. Kilmer, and D. P. O'Leary, *A framework for regularization via operator approximation*, SIAM J. Sci. Comput., 37 (2015), pp. B332–59.

[11] J. Chung, J. G. Nagy, and D. P. O'Leary, *A weighted GCV method for Lanczos hybrid regularization*, Electron. Trans. Numer. Anal., 28 (2008), pp. 149–167.

[12] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Comm. Pure Appl. Math., 63 (2010), pp. 1–38, https://doi.org/10.1002/cpa.20303.

[13] J. Demmel, L. Grigori, M. Hoemmen, and J. Langou, *Communication-optimal parallel and sequential QR and LU factorizations*, SIAM J. Sci. Comput., 34 (2012), pp. A206–A239.

[14] M. Donatelli, A. Neuman, and L. Reichel, *Square regularization matrices for large linear discrete ill-posed problems*, Numer. Linear Algebra Appl., 19 (2012), pp. 896–913.

[15] H. Dong, Y.and Garde and P. Hansen, *R3GMRES: Including prior information in GMRES-type methods for discrete inverse problems*, Electron. Trans. Numer. Anal., 42 (2014), pp. 136–146.

[16] M. Fornasier and H. Rauhut, *Compressive Sensing*, Springer New York, New York, NY, 2011, pp. 187–228, https://doi.org/10.1007/978-0-387-92920-0_6.

[17] S. Gazzola, P. Hansen, and J. G. Nagy., *IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 81 (2019), pp. 773–811, https://doi.org/10.1007/s11075-018-0570-7.

[18] S. Gazzola and J. Nagy, *Generalized Arnoldi-Tikhonov method for sparse reconstruction*, SIAM J. Sci. Comput., 36 (2014), https://doi.org/10.1137/130917673.

[19] S. Gazzola, P. Novati, and M. R. Russo, *On Krylov projection methods and Tikhonov regularization*, Electron. Trans. Numer. Anal., 44 (2015), pp. 83–123.

[20] S. Gazzola and M. Sabaté Landman, *Flexible GMRES for total variation regularization*, BIT Numerical Mathematics, (2019).

[21] S. Gazzola and M. Sabaté Landman, *Krylov Methods for Inverse Problems: surveying classical, and introducing new, algorithmic approaches*, GAMM-Mitt., (2020), https://doi.org/10.1002/GAMM.202000017.

[22] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[23] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1996.

[24] P. Hansen, Y. Dong, and K. Abe, *Hybrid enriched bidiagonalization for discrete ill-posed problems*, Numer. Linear Algebra Appl., 26 (2019), p. e2230, https://doi.org/10.1002/nla.2230.

[25] P. C. Hansen, *Discrete Inverse Problems*, SIAM, Philadelphia, 2010, https://doi.org/10.1137/1.9780898718836.

[26] P. C. Hansen and J. S. Jorgensen, *AIR Tools II: Algebraic Iterative Reconstruction Methods, Improved Implementation*, Numer. Algorithms, (2018), https://doi.org/10.1007/

s11075-017-0430-x.

[27] G. Huang, A. Lanza, S. Morigi, L. Reichel, and F. Sgallari., *Majorization–minimization generalized Krylov subspace methods for ℓ_p-ℓ_q optimization applied to image restoration*, BIT Numerical Mathematics, 57 (2017), pp. 351–378.

[28] Y. Huang and Z. Jia, *Some results on the regularization of LSQR for large-scale discrete ill-posed problems*, Sci. China Math., 60 (2017), pp. 701–718.

[29] M. E. Kilmer and E. de Sturler, *Recycling subspace information for diffuse optical tomography*, SIAM J. Sci. Comput., 27 (2006), pp. 2140–2166.

[30] M. E. Kilmer and D. P. O'Leary., *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221, https://doi.org/10.1137/S0895479899345960.

[31] A. Lanza, S. Morigi, L. Reichel, and F. Sgallari, *A generalized Krylov subspace method for ℓ_p-ℓ_q minimization*, SIAM J. Sci. Comput., 37 (2015).

[32] A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari, *Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization*, Numer. Math., 136 (2016), https://doi.org/10.1007/s00211-016-0842-x.

[33] X. Li, D. Sun, , and K. Toh, *A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems*, SIAM J. Optim., 28 (2018), pp. 433–458, https://doi.org/10.1137/16M1097572.

[34] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.

[35] J. Nagy, K. Palmer, and L. Perrone, *Iterative methods for image deblurring; a MATLAB object-oriented approach.*, Numer. Algorithms, 36 (2004), p. 73–93.

[36] J. G. Nagy and D. P. O'Leary, *Restoring images degraded by spatially-variant blur*, SIAM J. Sci. Comput., 19 (1998), pp. 1063–1082.

[37] D. P. O'Leary and J. A. Simmons, *A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 474–489.

[38] M. Parks, E. de Sturler, G. Mackey, D. Johnson, and S. Maiti, *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput., 28 (2004), https://doi.org/10.1137/040607277.

[39] P. Rodriguez and B. Wohlberg, *An efficient algorithm for sparse representations with ℓ^p data fidelity term*, In Proceedings of the 4th IEEE Andean Technical Conference (ANDESCON), (2008).

[40] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469, https://doi.org/10.1137/0914028.

[41] Y. Saad, *Analysis of augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 18 (1995), https://doi.org/10.1137/S0895479895294289.

[42] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, PA, USA, 2nd ed., 2003.

[43] V. Simoncini and D. Szyld, *Flexible inner-outer Krylov subspace methods*, SIAM J. Numer. Anal., 40 (2002), https://doi.org/10.1137/S0036142902401074.

[44] V. Simoncini and D. Szyld, *Recent computational developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59, https://doi.org/10.1002/nla.499.

[45] L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*, SIAM, Philadelphia, PA, 2017, https://doi.org/10.1137/1.9781611974928.

[46] S. J. Wright, R. D. Nowak, and M. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493, https://doi.org/10.1109/icassp.2008.4518374.